

Comment proposal for Nature Reviews Drug Discovery

eTRANSafe: Data science to empower translational safety assessment

List of authors for the printed version:

Ferran Sanz*, François Pognan*, Thomas Steger-Hartmann*, Carlos Díaz and the eTRANSafe Consortium**

* Corresponding authors

**See online version for the full list of authors

Full list of authors for the online version:

Ferran Sanz, François Pognan, Thomas Steger-Hartmann, Carlos Díaz, Shoji Asakura, Alexander Amberg, Nathalie Bécourt-Lhote, Niklas Blomberg, Nicolas Bosc, Katharine Briggs, Frank Bringezu, Claire Brulle-Wohlhueter, Søren Brunak, Ruud Bueters, Giulia Callegaro, Salvador Capella, Emilio Centeno, Javier Corvi, Mark T. D. Cronin, Philip Drew, Guillemette Duchateau-Nguyen, Gerhard F. Ecker, Sylvia Escher, Eloy Felix, Miguel Ferreira, Markus Frericks, Laura I. Furlong, Robert Geiger, Catherine George, Melanie Grandits, Dragomir Ivanov-Draganov, Jean Kilgour-Christie, Tevfik Kiziloren, Jan A. Kors, Naoki Koyama, Annika Kreuchwig, Andrew R. Leach, Miguel Angel Mayer, Peter Monecke, Wolfgang Muster, Chihiro Miyamoto Nakazawa, Gavin Nicholson, Rowan Parry, Manuel Pastor, Janet Piñero, Nils Oberhauser, Juan Manuel Ramírez-Anguita, Adrián Rodrigo, Aljosa Smajic, Markus Schaefer, Sebastian Schieferdecker, Inari Soininen, Emma Terricabras, Panuwat Trairatphisan, Sean C. Turner, Alfonso Valencia, Bob van de Water, Johan L. van der Lei, Erik M. van Mulligen, Esther Vock, David Wilkinson

Subject terms: Drug Safety. Toxicology. Data Sharing. Data Integration. Computational Platform. *In silico* Predictive Models.

Standfirst sentence: Thirteen pharmaceutical companies have shared and integrated preclinical and clinical data for creating computational resources that enhances translational drug safety assessment

Drug discovery and development is a knowledge intensive process and, consequently, it benefits from advances in data science and technologies, such as automated data curation, data integration and artificial intelligence. Particularly, data sharing beyond the boundaries of each pharmaceutical company offers interesting opportunities for accelerating and improving drug discovery and development. However, balancing collaboration and

confidentiality remains a complex challenge for the industry. Given that drug safety does not constitute the main playground for the competition between pharmaceutical companies, this is a domain in which it is more feasible to develop collaborative projects. Legacy drug safety data may be shared among companies with the aim of improving drug safety assessment. Significant steps forward in this direction were carried out by the eTOX project focused on pre-clinical toxicological data sharing, which was carried out from 2010 to 2017 (<http://www.etoxproject.eu>) (1). More recently, the eTRANSafe project addressing translational drug safety assessment through integrative knowledge management, has been running from 2017 to 2023 (<https://etransafe.eu/>) (2). Both projects are part of the Innovative Medicines Initiative (IMI) funded by the European Commission and the European Federation of Pharmaceutical Industries and Associations (EFPIA).

Achievements of the eTRANSafe initiative

Although both eTOX and eTRANSafe were grounded on the potential of legacy data sharing in drug safety assessment, eTRANSafe incorporates substantial advancements in comparison to the eTOX project. While only aggregated data were shared in eTOX, repeated dose toxicology (RDT) studies at the individual animal level of detail have been shared in eTRANSafe. These RDT data are coded in the CDISC Standard for Exchange of Nonclinical Data (SEND) and they are electronically extracted from the LIMS systems of the pharmaceutical companies, avoiding human transcription errors. Almost 10,000 RDT studies, not available anywhere, have been shared by the pharmaceutical companies in eTRANSafe. Non-SEND historical data, like those previously shared in eTOX, have been mapped to SEND to maximise standardized RDT data available in eTRANSafe. Another advancement in comparison to eTOX has been the extraction of treatment-related findings from the free-text RDT study reports by means of a natural language processing tool developed in the project, as well as a SEND-adapted 'Study Report (SR) Domain Template & Editor' created for incorporating this information into the eTRANSafe preclinical database. Moreover, pharmaceutical companies have also exchanged in vitro off-target safety pharmacology data and pharmacovigilance data extracted from periodic safety update reports (PSURs), which have been used to create new legacy databases. Currently, eTRANSafe integrates information from 12 databases (DB) that contain proprietary and publicly available data generated in preclinical and clinical research (eTRANSafe preclinical DB, eTOX DB, off-target pharmacology DB, ChEMBL, DB of drugs withdrawn due to safety concerns, DB of drug adversities extracted from clinicaltrials.gov (3), DB of biomarkers extracted by text mining from clinicaltrials.gov, DailyMed, DB on drug adverse events from publications in Medline, FAERS, FDA DILrank, and PSUR DB). The inclusion of clinical databases allowed significant opportunities for translational analysis, i.e., investigating the concordance between the results obtained in preclinical studies and the safety observations in clinical studies. The project has also developed new tools for exploring knowledge about the molecular basis of toxicological events (4).

A modular platform for data integration and exploitation

The eTRANSAFE project has developed a modular software platform (ToxHub) that allows the integrated exploration and exploitation of all the above-mentioned databases. ToxHub incorporates a software tool for a universal and user-friendly incorporation of future new databases, a computational service enabling preclinical-clinical semantic interoperability by aligning the preclinical terminologies of SEND with the clinical vocabulary of MedDRA, a chemistry service for identifying in an unambiguous way the structures of the drugs included in the different databases by means of SMILES representations and InChI keys, and a structural similarity service for expanding the queries with compounds showing structural similarity or containing given substructures. ToxHub also incorporates advanced software for data visualization (Sirona), finely tuned to support the needs of drug safety professionals, as well as a computational platform (Flame) (5) designed for a user-friendly development, documentation and use of predictive models based on machine learning algorithms. Moreover, Flame incorporates functionalities for the development of ensemble models and for the export of models without including any kind of information about the compounds of the training set used. This constitutes a critical functionality to enable the cooperative model building using non-sharable data. A collection of predictive models has already been developed and made available in the ToxHub, including a series of predictive models of drug-transporter interactions. Several software developments of eTRANSAFE are distributed as open source.

Application and impact in pharmaceutical research

The ToxHub platform is already operational in many of the pharmaceutical companies involved in eTRANSAFE, which are using the platform in their daily practice. Several use cases have been performed to demonstrate how ToxHub improves our ability to carry out preclinical vs. clinical drug safety contrasts, such as the comparative analysis of the skin toxicity of kinase inhibitors.

Kinase inhibitors are one of the most important drug classes in cancer treatment, but their inhibition of key processes of the cell cycle generates numerous adverse events, which are sometimes dose-limiting or even life-threatening. Skin reactions represent one of the dose-limiting toxicities of kinase inhibitors. Most of the adverse events are generally identified during preclinical development in animal studies. However, systematic investigations on how far these observations translate into serious adverse events and which animal species is the most predictive for specific events are scarce. ToxHub allows to access preclinical data in sufficient amount and granularity to perform powerful comparisons in both the preclinical and clinical domains. In this use case, preclinical data was searched in ToxHub for skin findings reported with kinase inhibitors. Preclinical data for more than one species was found for gefitinib, imatinib, nilotinib and vandetanib. In a second step, the Clinical Trials and the FAERS databases integrated into the ToxHub were searched for clinical skin-related findings in these four kinase inhibitors. Nilotinib had only few entries regarding severe skin reactions in Clinical Trials and no entries in FAERS, whereas gefitinib, imatinib and

vandetanib had various entries of severe skin findings such as ulcer and dermatitis. The comparison of the preclinical and the clinical results indicated that in this case, rat was the most sensitive species to detect severe skin findings, such as atrophy and ulcer also occurring clinically with gefitinib and imatinib, whereas non-human primates were evidently less sensitive. This higher concordance of rat compared to monkey regarding skin reactions of kinase inhibitors is in accordance with an analysis previously performed for a wider range of drugs (6).

In addition, eTRANSAFE has successfully prototyped a promising application, consisting in the reuse of legacy control groups' data obtained in preclinical in vivo studies under GLP conditions to develop Virtual Control Groups (7). These Virtual Control Groups could in time substitute part of the animals currently used in experimental control groups, significantly helping to reduce the number of animals used in preclinical safety assessment, in line with the 3Rs policies (Replacement, Reduction and Refinement).

In conclusion, the integration of preclinical and clinical data combined with a computational toolbox customized to the needs of pharmaceutical safety experts, enables efficient data retrieval and translational analysis, and exemplifies how modern data science can increase comprehensiveness, accuracy, and speed in assessing safety issues during drug development.

References

- 1) Sanz F *et al.* Legacy data sharing to improve drug safety assessment: the eTOX project. *Nat. Rev. Drug Discov.* **16**, 811-812 (2017).
- 2) Pognan F *et al.* The eTRANSAFE Project on Translational Safety Assessment through Integrative Knowledge Management: Achievements and Perspectives. *Pharmaceuticals (Basel)*. **14**, 237 (2021).
- 3) Piñero J *et al.* Genomic and Proteomic Biomarker Landscape in Clinical Trials. *Comput. Struct. Biotechnol. J.* In press.
- 4) Callegaro G *et al.* The human hepatocyte TXG-MAPr: gene co-expression network modules to support mechanism-based risk assessment. *Arch. Toxicol.* **95**, 3745-3775 (2021).
- 5) Pastor M *et al.* Flame: an open-source framework for model development, hosting, and usage in production environments. *J. Cheminform.* **13**, 31 (2021).
- 6) Clark M *et al.* A big data approach to the concordance of the toxicity of pharmaceuticals in animals and humans. *Regul. Toxicol. Pharmacol.* **96**, 94-105 (2018).
- 7) Steger-Hartmann T *et al.* Introducing the concept of virtual control groups into preclinical toxicology testing. *ALTEX.* **37**, 343-349 (2020).

Competing interests statement

Some of the authors are employed in the pharmaceutical industry or in SMEs, as indicated in their affiliations.

Affiliations of the authors

Ferran Sanz, Emilio Centeno, Miguel Angel Mayer, Manuel Pastor, Juan Manuel Ramírez-Anguila, Emma Terricabras

Hospital del Mar Medical Research Institute (IMIM), Universitat Pompeu Fabra, 08003 Barcelona, Spain

François Pognan, Jean Kilgour-Christie, Nils Oberhauser

Novartis Pharma AG, CH-4056 Basel, Switzerland

Thomas Steger-Hartmann, Annika Kreuchwig

Bayer AG, Pharmaceuticals, 13353 Berlin, Germany

Carlos Díaz, Inari Soininen

Synapse Research Management Partners, 28006 Madrid, Spain

Shoji Asakura, Naoki Koyama, Chihiro Miyamoto Nakazawa

Eisai Co., Ltd. Tsukuba, Ibaraki, Japan

Alexander Amberg, Claire Brulle-Wohlhueter, Peter Monecke, Markus Schaefer

Preclinical Safety, Sanofi, 65926 Frankfurt am Main, Germany

Nathalie Bécourt-Lhote

Non-Clinical Drug Safety, Translational Medicine, Servier, 92284 Suresnes, France

Niklas Blomberg

ELIXIR, Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK

Nicolas Bosc, Eloy Felix, Tevfik Kiziloren, Andrew R. Leach

European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK

Katherine Briggs, David Wilkinson

Lhasa Limited, Leeds LS11 5PS, UK

Frank Bringezu

Chemical and Preclinical Safety, Merck Healthcare KGaA, 64293 Darmstadt, Germany

Søren Brunak

Technical University of Denmark (DTU), 2800 Lyngby, Denmark

(Current address: Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, 2200 Copenhagen, Denmark)

Ruud Bueters

Janssen Research & Development, Janssen Pharmaceutica N.V., B-2340 Beerse, Belgium

Giulia Callegaro, Bob van de Water

Leiden Academic Centre for Drug Research (LACDR), Leiden University, 2300 RA Leiden, The Netherlands

Salvador Capella, Javier Corvi, Jordi Alfonso Valencia

Barcelona Supercomputing Center, 08034 Barcelona, Spain

Mark T. D. Cronin

Liverpool John Moores University, Liverpool L3 3AF, UK

Philip Drew

PDS Consultants, Leicester LE1 5XY, UK

Guillemette Duchateau-Nguyen, Dragomir Ivanov-Draganov, Wolfgang Muster

Roche Pharmaceutical Research & Early Development, Pharmaceutical Sciences, Roche Innovation Center, 4070 Basel, Switzerland

Gerhard F. Ecker, Melanie Grandits, Aljosa Smajic

Universität Wien, 1090 Vienna, Austria

Sylvia Escher

Fraunhofer Institute for Toxicology and Experimental Medicine (ITEM), 30625 Hannover, Germany

Miguel Ferreira, Adrián Rodrigo

GMV, 28760 Tres Cantos, Madrid, Spain

Markus Frericks, Robert Geiger

BASF, 67056 Ludwigshafen, Germany

Laura I. Furlong, Janet Piñero

MedBioinformatics Solutions SL, 08018 Barcelona, Spain

Hospital del Mar Medical Research Institute (IMIM), Universitat Pompeu Fabra, 08003 Barcelona, Spain

Catherine George

Ipsen Innovation, 91940 Les Ulis, France

Jan A. Kors, Rowan Parry, Erik M. van Mulligen

Department of Medical Informatics, Erasmus University Medical Center, 3015 GD Rotterdam, the Netherlands

Gavin Nicholson

Optibrium Ltd, Cambridge Innovation Park, Cambridge CB25 9GL, UK

Johan L. van der Lei

Department of Medical Informatics, Erasmus University Medical Center, 3015 GD
Rotterdam, The Netherlands

Synapse Research Management Partners, 28006 Madrid, Spain

Esther Vock, Sebastian Schieferdecker

Boehringer Ingelheim Pharma GmbH & Co. KG, 88379 Biberach an der Riss, Germany

Panuwat Trairatphisan, Sean C. Turner

AbbVie Germany GmbH & Co. KG, 67061 Ludwigshafen am Rhein, Germany